JOURNAL OF CHEMOMETRICS
J. Chemometrics 2001; **15**: 439–453
DOI: 10.1002/cem.633

On the self-weighted alternating trilinear decomposition algorithm—the property of being insensitive to excess factors used in calculation

Zeng-Ping Chen, Hai-Long Wu and Ru-Qin Yu*

College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, People's Republic of China

SUMMARY

PARAFAC is well known as an iterative trilinear decomposition method. In practice, an accurate estimation of the number of underlying factors is required; otherwise it is difficult to guarantee the chemical meaning of the results obtained. The absence of a versatile chemometric method for factor estimation has often caused problems for analysts. With the advent of ATLD followed by SWATLD, the above relatively strict constraint can be relaxed. Experiments have shown that the profiles of the underlying factors can be extracted by ATLD (Wu *et al.*, *J. Chemometrics* 1998; **12**: 1) and SWATLD (Chen *et al.*, *Chemometrics Intell. Lab. Syst*. 2000; 52: 75) as long as the number of factors used in calculation is no less than the number of actual factors. In other words, an overestimation of the number of factors will not affect the obtainment of chemically meaningful results by SWATLD and ATLD. In this paper the authors provide some simple mathematical explanations of this valuable property of SWATLD. Along with these explanations, some other properties of SWATLD as well as some guidelines for designing new trilinear decomposition methods are also discussed. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS: iterative trilinear decomposition methods; PARAFAC; ATLD; SWATLD

1. INTRODUCTION

Trilinear decomposition methods have the very important advantage of being capable of determining the contents of sought-for analytes in a mixture in spite of the existence of unknown interferences or background [1]. Attracted by this property, more and more scientists have been engaging in the research of designing new algorithms and their potential applications [2–7]. Trilinear decomposition methods can be loosely classified into non-iterative algorithms, represented by GRAM [2], and iterative ones, exemplified by PARAFAC [3,4]. The theory and properties of GRAM have been documented elsewhere [8–10]. In this paper the focus is on iterative trilinear decomposition algorithms. Among the family of iterative algorithms, PARAFAC, first introduced in psychology, is the most famous and widely accepted one in chemometrics, its theory having been thoroughly

* Correspondence to: Ru-Qin Yu, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, People's Republic of China.
E-mail: rqyu@mail.hunu.edu.cn
Contract/grant sponsor: National Natural Science Foundation of China; Contract/grant number: 29735150

Received 28 September 1999
Accepted 30 March 2000

investigated [3,4,11,12]. Although PARAFAC has been successfully applied to many chemical problems [13–16], the requirement of an accurate estimation of the number of underlying factors is an obstacle to its wider applications, since there is no such method which can guarantee an accurate estimation of the number of underlying factors under all circumstances.

Recently, Wu *et al.* proposed an ATLD algorithm [17]. Their algorithm is characterized by the property of being insensitive to excess factors used in calculation. In other words, the final results of ATLD will contain the profiles of the actual factors, provided that the number of factors used in calculation is no less than the number of actual factors. The significance of such a feature is obvious, for it can release analysts from the troublesome chemical rank estimation tasks. Unfortunately, there is little explanation of this feature in the original paper.

The present authors have also designed an iterative trilinear decomposition algorithm, called SWATLD [18], which possesses the same attractive feature as ATLD. In this paper we try to give some mathematical explanations of this feature. In addition, some other properties of SWATLD and some guidelines for designing new trilinear decomposition algorithms are also discussed.

2. MODEL AND ALGORITHM

2.1. Nomenclature

Throughout this paper, scalars are represented by lower-case italics, vectors are denoted by bold lower-case characters, bold capitals designate two-way matrices and underlined bold capitals symbolize three-way arrays. Before reading the main text of this paper, readers are recommended to refer to Appendix I for detailed nomenclature information.

2.2. The model

In second-order linear calibration the famous trilinear decomposition model proposed by Harshman [3] and Carroll and Chang [4] has been widely accepted owing to its consistency with Beer's law in chemistry. Assuming there is no noise, every element x_{ijk} of data array $\underline{\mathbf{X}}$ can be expressed as the sum of a series of products:

$$x_{ijk} = \sum_{f=1}^{F} a_{if}^{\circ} b_{if}^{\circ} c_{kf}^{\circ}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K \tag{1}$$

Provided that $k_1 + k_2 + k_3 \geq 2F + 2$ (k_1, k_2 and k_3 are the k-ranks [19] of the three underlying loading matrices in data array $\underline{\mathbf{X}}$ respectively) [20,21], the above decomposition is unique up to some scaling and permutation indeterminacy.

Along with the model, Harshman [3] and Carroll and Chang [4] proposed an alternating least squares approach to solve the above decomposition problem by successively assuming the loading matrices in two modes to be known and estimating the unknown parameters of the third mode. It has been pointed out that only when the number of underlying factors has been correctly estimated can uniqueness be guaranteed for the results obtained by the alternating least squares algorithm. A wrong choice of the number of factors used in calculation generally leads to erroneous results. In practice, the absence of a versatile method for component estimation limits the application of this algorithm. Such a situation calls for new algorithms with mild constraints. As far as we know, ATLD of Wu *et al.* [17] is the only iterative algorithm published so far with the property of being insensitive to excess factors used in calculation. Unfortunately, there is little explanation of this feature in their original paper. Recently, the present authors have also proposed an algorithm, called self-weighted alternating

trilinear decomposition (SWATLD) [18]. It also possesses this attractive property, which will be explained with more or less mathematical rigour in the following sections.

2.3. *Self-weighted alternating trilinear decomposition (SWATLD)*

Before the explanations of the property, a brief description of SWATLD is necessary. In fact, SWATLD aims to alternately minimize the following three objective functions (i.e. assuming **A** and **B** are known, minimize $S(\mathbf{C})$ to obtain loading matrix **C**; with **A** and the new **C**, minimize $S(\mathbf{B})$ to renew **B**; and then minimize $S(\mathbf{A})$ with the latest **B** and **C** to obtain a new **A**):

$$S(\mathbf{C}) = \sum_{k=1}^{K} (\|(\mathbf{X}_{..k}^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}})^{+} - \mathbf{B}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}}))\mathrm{diag}(\mathrm{sqrt}(\mathbf{1}./\mathrm{diagm}(\mathbf{B}^{\mathrm{T}}\mathbf{B})))\|_{\mathrm{F}}^2 + \|(\mathbf{X}_{..k}(\mathbf{B}^{\mathrm{T}})^{+} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}}))\mathrm{diag}(\mathrm{sqrt}(\mathbf{1}./\mathrm{diagm}(\mathbf{A}^{\mathrm{T}}\mathbf{A})))\|_{\mathrm{F}}^2) \quad (2)$$

$$S(\mathbf{B}) = \sum_{j=1}^{J} (\|(\mathbf{X}_{j.}^{\mathrm{T}}(\mathbf{C}^{\mathrm{T}})^{+} - \mathbf{A}\mathrm{diag}(\mathbf{b}_j^{\mathrm{T}}))\mathrm{diag}(\mathrm{sqrt}(\mathbf{1}./\mathrm{diagm}(\mathbf{A}^{\mathrm{T}}\mathbf{A})))\|_{\mathrm{F}}^2 + \|(\mathbf{X}_{j.}(\mathbf{A}^{\mathrm{T}})^{+} - \mathbf{C}\mathrm{diag}(\mathbf{b}_j^{\mathrm{T}}))\mathrm{diag}(\mathrm{sqrt}(\mathbf{1}./\mathrm{diagm}(\mathbf{C}^{\mathrm{T}}\mathbf{C})))\|_{\mathrm{F}}^2) \quad (3)$$

$$S(\mathbf{A}) = \sum_{i=1}^{I} (\|(\mathbf{X}_{i..}^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}})^{+} - \mathbf{C}\mathrm{diag}(\mathbf{a}_i^{\mathrm{T}}))\mathrm{diag}(\mathrm{sqrt}(\mathbf{1}./\mathrm{diagm}(\mathbf{C}^{\mathrm{T}}\mathbf{C})))\|_{\mathrm{F}}^2 + \|(\mathbf{X}_{i..}(\mathbf{C}^{\mathrm{T}})^{+} - \mathbf{B}\mathrm{diag}(\mathbf{a}_i^{\mathrm{T}}))\mathrm{diag}(\mathrm{sqrt}(\mathbf{1}./\mathrm{diagm}(\mathbf{B}^{\mathrm{T}}\mathbf{B})))\|_{\mathrm{F}}^2) \quad (4)$$

Here $\|\cdot\|_{\mathrm{F}}$ represents the Frobenius matrix norm; ./ denotes array division, e.g. suppose $\mathbf{x} = (x_i)$ and $\mathbf{y} = (y_i)$, then $\mathbf{x./y} = (x_i/y_i)$; sqrt(·) is a square root operator; and **1** is a vector of length N with all elements equal to one. The three diagonal matrices diag(sqrt(**1**./diagm($\mathbf{A}^{\mathrm{T}}\mathbf{A}$))), diag(sqrt(**1**./diagm($\mathbf{B}^{\mathrm{T}}\mathbf{B}$))) and diag(sqrt(**1**./diag($\mathbf{C}^{\mathrm{T}}\mathbf{C}$))) function as weight matrices to balance the two parts of each objective function.

Based on objective functions (2)–(4), SWATLD was designed as follows.

1. Randomly initialize loading matrices **A** and **B**.

2. $\mathbf{c}_k^{\mathrm{T}} = \frac{1}{2}(\mathrm{diagm}(\mathbf{B}^{+}\mathbf{X}_{..k}^{\mathrm{T}}\mathbf{A})./\mathrm{diagm}(\mathbf{A}^{\mathrm{T}}\mathbf{A}) + \mathrm{diagm}(\mathbf{A}^{+}\mathbf{X}_{..k}\mathbf{B})./\mathrm{diagm}(\mathbf{B}^{\mathrm{T}}\mathbf{B})), \quad k = 1, \ldots, K$ (5)

3. $\mathbf{b}_j^{\mathrm{T}} = \frac{1}{2}(\mathrm{diagm}(\mathbf{A}^{+}\mathbf{X}_{j.}^{\mathrm{T}}\mathbf{C})./\mathrm{diagm}(\mathbf{C}^{\mathrm{T}}\mathbf{C}) + \mathrm{diagm}(\mathbf{C}^{+}\mathbf{X}_{j.}\mathbf{A})./\mathrm{diagm}(\mathbf{A}^{\mathrm{T}}\mathbf{A})), \quad j = 1, \ldots, J$ (6)

4. $\mathbf{a}_i^{\mathrm{T}} = \frac{1}{2}(\mathrm{diagm}(\mathbf{C}^{+}\mathbf{X}_{i..}^{\mathrm{T}}\mathbf{B})./\mathrm{diagm}(\mathbf{B}^{\mathrm{T}}\mathbf{B}) + \mathrm{diagm}(\mathbf{B}^{+}\mathbf{X}_{i..}\mathbf{C})./\mathrm{diagm}(\mathbf{C}^{\mathrm{T}}\mathbf{C})), \quad i = 1, \ldots, I$ (7)

5. Update **C, B** and **A** according to steps 2–4 until a certain stop criterion has been reached.

Like ATLD, our experiments showed that SWATLD is also insensitive to excess factors used in computation. It can extract the actual profiles of underlying factors even if $N > F$.

3. EXPLANATIONS OF THE PROPERTY OF SWATLD BEING INSENSITIVE TO EXCESS FACTORS USED IN CALCULATION

All the theorems and explanations in the following are based on the assumption of absence of noise.

Theorem 1

If the ranks of the underlying loading matrices $\mathring{\mathbf{A}}$, $\mathring{\mathbf{B}}$ and $\mathring{\mathbf{C}}$ are equal, say F, and no noise exists, then there are matrices **A, B** and **C** with N columns ($N \geq F$) which satisfy $S(\mathbf{C}) = 0$, $S(\mathbf{B}) = 0$ and $S(\mathbf{A}) = 0$.

Proof. (a) When $N = F$. Assuming $\mathbf{A} = \mathring{\mathbf{A}}$, $\mathbf{B} = \mathring{\mathbf{B}}$ and $\mathbf{C} = \mathring{\mathbf{C}}$, since $\text{rank}(\mathring{\mathbf{A}}) = \text{rank}(\mathring{\mathbf{B}}) = \text{rank}(\mathring{\mathbf{C}}) = F$, the following equations hold:

$$\mathbf{X}_{..k}^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}})^{+} = \mathring{\mathbf{B}}\text{diag}(\mathring{\mathbf{c}}_k^{\mathrm{T}})\mathring{\mathbf{A}}^{\mathrm{T}}(\mathring{\mathbf{A}}^{\mathrm{T}})^{+} = \mathring{\mathbf{B}}\text{diag}(\mathring{\mathbf{c}}_k^{\mathrm{T}}) = \mathbf{B}\text{diag}(\mathbf{c}_k^{\mathrm{T}}), \qquad k = 1, 2, \ldots, K$$

$$\mathbf{X}_{..k}(\mathbf{B}^{\mathrm{T}})^{+} = \mathring{\mathbf{A}}\text{diag}(\mathring{\mathbf{c}}_k^{\mathrm{T}})\mathring{\mathbf{B}}^{\mathrm{T}}(\mathring{\mathbf{B}}^{\mathrm{T}})^{+} = \mathring{\mathbf{A}}\text{diag}(\mathring{\mathbf{c}}_k^{\mathrm{T}}) = \mathbf{A}\text{diag}(\mathbf{c}_k^{\mathrm{T}}), \quad k = 1, 2, \ldots, K$$

Consequently, $S(\mathbf{C}) = 0$. Similarly, it can be demonstrated that $S(\mathbf{A}) = 0$ and $S(\mathbf{B}) = 0$.

(b) When $N > F$. For simplicity of description it is assumed that $F = 4$ and $N = 6$ (note that the values of F and N are set randomly; they will not affect the generality of the following conclusions).

$$\mathring{\mathbf{A}} = \left(\mathring{\mathbf{a}}_1, \mathring{\mathbf{a}}_2, \mathring{\mathbf{a}}_3, \mathring{\mathbf{a}}_4\right), \qquad \mathring{\mathbf{B}} = \left(\mathring{\mathbf{b}}_1, \mathring{\mathbf{b}}_2, \mathring{\mathbf{b}}_3, \mathring{\mathbf{b}}_4\right), \qquad \mathring{\mathbf{C}} = \left(\mathring{\mathbf{c}}_1, \mathring{\mathbf{c}}_2, \mathring{\mathbf{c}}_3, \mathring{\mathbf{c}}_4\right)$$

Suppose

$$\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2), \qquad \mathbf{A}_1 = \mathring{\mathbf{A}}, \qquad \mathbf{A}_2 = (\mathring{\mathbf{a}}_1, \mathring{\mathbf{a}}_2)$$

$$\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2), \qquad \mathbf{B}_1 = \mathring{\mathbf{B}}, \qquad \mathbf{B}_2 = \left(\mathring{\mathbf{b}}_1, \mathring{\mathbf{b}}_2\right)$$

$$\mathbf{C} = \left(\tfrac{1}{2}\mathring{\mathbf{c}}_1, \tfrac{1}{2}\mathring{\mathbf{c}}_2, \mathring{\mathbf{c}}_3, \mathring{\mathbf{c}}_4, \tfrac{1}{2}\mathring{\mathbf{c}}_1, \tfrac{1}{2}\mathring{\mathbf{c}}_2\right)$$

According to Reference [22],

$$\mathbf{A}^{+} = \begin{pmatrix} \mathbf{A}_1^{+} - \mathbf{A}_1^{+}\mathbf{A}_2\mathbf{K}_{\mathbf{A}}^{-1}\mathbf{A}_2^{\mathrm{T}}(\mathbf{A}_1^{+})^{\mathrm{T}}\mathbf{A}_1^{+} \\ \mathbf{K}_{\mathbf{A}}^{-1}\mathbf{A}_2^{\mathrm{T}}(\mathbf{A}_1^{+})^{\mathrm{T}}\mathbf{A}_1^{+} \end{pmatrix}, \qquad \mathbf{B}^{+} = \begin{pmatrix} \mathbf{B}_1^{+} - \mathbf{B}_1^{+}\mathbf{B}_2\mathbf{K}_{\mathbf{B}}^{-1}\mathbf{B}_2^{\mathrm{T}}(\mathbf{B}_1^{+})^{\mathrm{T}}\mathbf{B}_1^{+} \\ \mathbf{K}_{\mathbf{B}}^{-1}\mathbf{B}_2^{\mathrm{T}}(\mathbf{B}_1^{+})^{\mathrm{T}}\mathbf{B}_1^{+} \end{pmatrix}$$

$$\mathbf{K}_{\mathbf{A}} = \mathbf{I} + \mathbf{A}_2^{\mathrm{T}}(\mathbf{A}_1^{+})^{\mathrm{T}}\mathbf{A}_1^{+}\mathbf{A}_2, \qquad \mathbf{K}_{\mathbf{B}} = \mathbf{I} + \mathbf{B}_2^{\mathrm{T}}(\mathbf{B}_1^{+})^{\mathrm{T}}\mathbf{B}_1^{+}\mathbf{B}_2$$

Suppose

$$\mathbf{A}_1{}^{+} = \begin{pmatrix} \mathbf{pa}_1^{\mathrm{T}} \\ \mathbf{pa}_2^{\mathrm{T}} \\ \mathbf{pa}_3^{\mathrm{T}} \\ \mathbf{pa}_4^{\mathrm{T}} \end{pmatrix}, \qquad \mathbf{B}_1{}^{+} = \begin{pmatrix} \mathbf{pb}_1^{\mathrm{T}} \\ \mathbf{pb}_2^{\mathrm{T}} \\ \mathbf{pb}_3^{\mathrm{T}} \\ \mathbf{pb}_4^{\mathrm{T}} \end{pmatrix}$$

Then

$$\mathbf{A}^{+} = \begin{pmatrix} \frac{1}{2}\mathbf{pa}_1^{\mathrm{T}} \\ \frac{1}{2}\mathbf{pa}_2^{\mathrm{T}} \\ \mathbf{pa}_3^{\mathrm{T}} \\ \mathbf{pa}_4^{\mathrm{T}} \\ \frac{1}{2}\mathbf{pa}_1^{\mathrm{T}} \\ \frac{1}{2}\mathbf{pa}_2^{\mathrm{T}} \end{pmatrix}, \qquad \mathbf{B}^{+} = \begin{pmatrix} \frac{1}{2}\mathbf{pb}_1^{\mathrm{T}} \\ \frac{1}{2}\mathbf{pb}_2^{\mathrm{T}} \\ \mathbf{pb}_3^{\mathrm{T}} \\ \mathbf{pb}_4^{\mathrm{T}} \\ \frac{1}{2}\mathbf{pb}_1^{\mathrm{T}} \\ \frac{1}{2}\mathbf{pb}_2^{\mathrm{T}} \end{pmatrix}$$

After some simple calculations one has

$$\begin{aligned} S_1(\mathbf{C}) &= \sum_{k=1}^{K} \|(\mathbf{X}_{..k}^{\mathrm{T}}(\mathbf{A}^{+})^{\mathrm{T}} - \mathbf{B}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}}))\mathrm{diag}(\mathrm{sqrt}(\mathbf{1}./\mathrm{diagm}(\mathbf{B}^{\mathrm{T}}\mathbf{B})))\|_{\mathrm{F}}^2 \\ &= \sum_{k=1}^{K} \|(\overset{\circ}{\mathbf{B}}\mathrm{diag}(\overset{\circ}{\mathbf{c}}_k^{\mathrm{T}})\overset{\circ}{\mathbf{A}}{}^{\mathrm{T}}(\mathbf{A}^{+})^{\mathrm{T}} - \mathbf{B}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}}))\mathrm{diag}(\mathrm{sqrt}(\mathbf{1}./\mathrm{diagm}(\mathbf{B}^{\mathrm{T}}\mathbf{B})))\|_{\mathrm{F}}^2 \\ &= 0 \end{aligned}$$

$$\begin{aligned} S_2(\mathbf{C}) &= \sum_{k=1}^{K} \|(\mathbf{X}_{..k}(\mathbf{B}^{+})^{\mathrm{T}} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}}))\mathrm{diag}(\mathrm{sqrt}(\mathbf{1}./\mathrm{diagm}(\mathbf{A}^{\mathrm{T}}\mathbf{A})))\|_{\mathrm{F}}^2 \\ &= \sum_{k=1}^{K} \|(\overset{\circ}{\mathbf{A}}\mathrm{diag}(\overset{\circ}{\mathbf{c}}_k^{\mathrm{T}})\overset{\circ}{\mathbf{B}}{}^{\mathrm{T}}(\mathbf{B}^{+})^{\mathrm{T}} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}}))\mathrm{diag}(\mathrm{sqrt}(\mathbf{1}./\mathrm{diagm}(\mathbf{A}^{\mathrm{T}}\mathbf{A})))\|_{\mathrm{F}}^2 \\ &= 0 \end{aligned}$$

Therefore $S(\mathbf{C}) = S_1(\mathbf{C}) + S_2(\mathbf{C}) = 0$. Similarly, it can also be easily verified that $S(\mathbf{B}) = 0$ and $S(\mathbf{A}) = 0$.

This means that there exist solutions for loading matrixes **A, B** and **C** with N columns ($N \geq F$) which satisfy the three equations $S(\mathbf{C}) = 0$, $S(\mathbf{B}) = 0$ and $S(\mathbf{A}) = 0$.

Theorem 2

The columns of the loading matrices **A, B** and **C** obtained by SWATLD are linear combinations of the columns of $\overset{\circ}{\mathbf{A}}$, $\overset{\circ}{\mathbf{B}}$ and $\overset{\circ}{\mathbf{C}}$ respectively.

Proof. Through some simple transformations, Equations (5)–(7) can be rewritten as

$$\mathbf{C} = \frac{1}{2}\left(\sum_{i=1}^{I} \mathbf{X}_{i..}^{\mathrm{T}}\mathbf{B}\mathrm{diag}(\mathbf{pat}_i)\mathrm{diag}(\mathbf{1}./\mathrm{diagm}(\mathbf{B}^{\mathrm{T}}\mathbf{B})) + \sum_{j=1}^{J} \mathbf{X}_{.j.}\mathbf{A}\mathrm{diag}(\mathbf{pbt}_j)\mathrm{diag}(\mathbf{1}./\mathrm{diagm}(\mathbf{A}^{\mathrm{T}}\mathbf{A}))\right)$$

$$\mathbf{B} = \frac{1}{2}\left(\sum_{k=1}^{K} \mathbf{X}_{..k}^{\mathrm{T}}\mathbf{A}\mathrm{diag}(\mathbf{pct}_k)\mathrm{diag}(\mathbf{1}./\mathrm{diagm}(\mathbf{A}^{\mathrm{T}}\mathbf{A})) + \sum_{i=1}^{I} \mathbf{X}_{i..}\mathbf{C}\mathrm{diag}(\mathbf{pat}_i)\mathrm{diag}(\mathbf{1}./\mathrm{diagm}(\mathbf{C}^{\mathrm{T}}\mathbf{C}))\right)$$

$$\mathbf{A} = \frac{1}{2}\left(\sum_{j=1}^{J} \mathbf{X}_{.j.}^{\mathrm{T}}\mathbf{C}\mathrm{diag}(\mathbf{pbt}_j)\mathrm{diag}(\mathbf{1}./\mathrm{diagm}(\mathbf{C}^{\mathrm{T}}\mathbf{C})) + \sum_{k=1}^{K} \mathbf{X}_{..k}\mathbf{B}\mathrm{diag}(\mathbf{pct}_k)\mathrm{diag}(\mathbf{1}./\mathrm{diagm}(\mathbf{B}^{\mathrm{T}}\mathbf{B}))\right)$$

where $\mathbf{pct}_k$ is the kth row of matrix $(\mathbf{C}^{+})^{\mathrm{T}}$, $\mathbf{pat}_i$ is the ith row of matrix $(\mathbf{A}^{+})^{\mathrm{T}}$, and $\mathbf{pbt}_j$ is the jth row of matrix $(\mathbf{B}^{+})^{\mathrm{T}}$.

Therefore the columns of loading matrix **A** are linear combinations of the columns of matrices $\mathbf{X}_{.j.}^{\mathrm{T}}(j = 1, 2, \ldots, J)$ and $\mathbf{X}_{..k}$ (k = 1,2, …, K), the columns of loading matrix **B** are linear combinations of the columns of matrices $\mathbf{X}_{..k}^{\mathrm{T}}(k = 1, 2, \ldots, K)$ and $\mathbf{X}_{i..}$ (i = 1, 2, …, I), and the columns of loading matrix **C** are linear combinations of the columns of matrices $\mathbf{X}_{i..}^{\mathrm{T}}(i = 1, 2, \ldots, I)$ and $\mathbf{X}_{.j.}$ (j = 1, 2, …, J). As defined in Appendix I, $\mathbf{X}_{..k} = \mathring{\mathbf{A}}\mathrm{diag}(\mathring{\mathbf{c}}_k^{\mathrm{T}})\mathring{\mathbf{B}}^{\mathrm{T}}$ (k = 1, 2,…, K), $\mathbf{X}_{.j.} = \mathring{\mathbf{C}}\mathrm{diag}(\mathring{\mathbf{b}}_j^{\mathrm{T}})\mathring{\mathbf{A}}^{\mathrm{T}}$ (j = 1, 2,…, J) and $\mathbf{X}_{i..} = \mathring{\mathbf{B}}\mathrm{diag}(\mathring{\mathbf{a}}_I^{\mathrm{T}})\,\mathring{\mathbf{C}}^{\mathrm{T}}$ (i = 1, 2,…, I), so the columns of loading matrices **A**, **B** and **C** are linear combinations of the columns of the underlying loading matrices $\mathring{\mathbf{A}}$, $\mathring{\mathbf{B}}$ and $\mathring{\mathbf{C}}$ respectively.

Theorem 3

If **A, B** and **C** all have rank F, then $\mathbf{X}_{..k} = \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}(k = 1, 2, \ldots, K)$, $\mathbf{X}_{.j.} = \mathbf{C}\mathrm{diag}(\mathbf{b}_j^{\mathrm{T}})\mathbf{A}^{\mathrm{T}}(j = 1, 2, \ldots, J)$ and $\mathbf{X}_{i..} = \mathbf{B}\mathrm{diag}(\mathbf{a}_i^{\mathrm{T}})\mathbf{C}^{\mathrm{T}}(i = 1, 2, \ldots, I)$ when $S(\mathbf{C}) = 0$, $S(\mathbf{B}) = 0$ and $S(\mathbf{A}) = 0$.

Proof. When $S(\mathbf{C}) = 0$, $S(\mathbf{B}) = 0$ and $S(\mathbf{A}) = 0$, the following equations hold:

$$\begin{aligned}
&\mathbf{A}^{+}\mathbf{X}_{..k} - \mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}} = 0, \quad k = 1, 2, \ldots, K \\
&\mathbf{X}_{..k}(\mathbf{B}^{+})^{\mathrm{T}} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}}) = 0, \quad k = 1, 2, \ldots, K \\
&\mathbf{C}^{+}\mathbf{X}_{.j.} - \mathrm{diag}(\mathbf{b}_j^{\mathrm{T}})\mathbf{A}^{\mathrm{T}} = 0, \quad j = 1, 2, \ldots, J \\
&\mathbf{X}_{.j.}(\mathbf{A}^{+})^{\mathrm{T}} - \mathbf{C}\mathrm{diag}(\mathbf{b}_j^{\mathrm{T}}) = 0, \quad j = 1, 2, \ldots, J \\
&\mathbf{B}^{+}\mathbf{X}_{i..} - \mathrm{diag}(\mathbf{a}_i^{\mathrm{T}})\mathbf{C}^{\mathrm{T}} = 0, \quad i = 1, 2, \ldots, I \\
&\mathbf{X}_{i..}(\mathbf{C}^{+})^{\mathrm{T}} - \mathbf{B}\mathrm{diag}(\mathbf{a}_i^{\mathrm{T}}) = 0, \quad i = 1, 2, \ldots, I
\end{aligned} \tag{8}$$

Then

$$\mathbf{A}\mathbf{A}^{+}(\mathbf{X}_{..k} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}) = 0, \quad k = 1, 2, \ldots, K$$

$$(\mathbf{X}_{..k} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}})(\mathbf{B}^{+})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} = 0, \quad k = 1, 2, \ldots, K$$

$$\mathbf{C}\mathbf{C}^{+}(\mathbf{X}_{.j.} - \mathbf{C}\mathrm{diag}(\mathbf{b}_j^{\mathrm{T}})\mathbf{A}^{\mathrm{T}}) = 0, \quad j = 1, 2, \ldots, J$$

$$(\mathbf{X}_{j.} - \mathbf{C}\mathrm{diag}(\mathbf{b}_j^{\mathrm{T}})\mathbf{A}^{\mathrm{T}})(\mathbf{A}^{+})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} = 0, \quad j = 1, 2, \ldots, J$$

$$\mathbf{B}\mathbf{B}^{+}(\mathbf{X}_{i..} - \mathbf{B}\mathrm{diag}(\mathbf{a}_i^{\mathrm{T}})\mathbf{C}^{\mathrm{T}}) = 0, \quad i = 1, 2, \ldots, I$$

$$(\mathbf{X}_{i..} - \mathbf{B}\mathrm{diag}(\mathbf{a}_i^{\mathrm{T}})\mathbf{C}^{\mathrm{T}})(\mathbf{C}^{+})^{\mathrm{T}}\mathbf{C}^{\mathrm{T}} = 0, \quad i = 1, 2, \ldots, I$$

Owing to Theorem 2, one has

$$\mathbf{R}(\mathbf{X}_{..k} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}) \subseteq \mathbf{R}(\mathbf{A}) = \mathbf{R}((\mathbf{A}^{+})^{\mathrm{T}})$$

$$\mathbf{R}(\mathbf{X}_{.j.} - \mathbf{C}\mathrm{diag}(\mathbf{b}_j^{\mathrm{T}})\mathbf{A}^{\mathrm{T}}) \subseteq \mathbf{R}(\mathbf{C}) = \mathbf{R}((\mathbf{C}^{+})^{\mathrm{T}})$$

$$\mathbf{R}(\mathbf{X}_{i..} - \mathbf{B}\mathrm{diag}(\mathbf{a}_i^{\mathrm{T}})\mathbf{C}^{\mathrm{T}}) \subseteq \mathbf{R}(\mathbf{B}) = \mathbf{R}((\mathbf{B}^{+})^{\mathrm{T}})$$

Therefore

$$\begin{aligned} \mathbf{X}_{..k} &= \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}, \quad k = 1, 2, \ldots, K \\ \mathbf{X}_{.j.} &= \mathbf{C}\mathrm{diag}(\mathbf{b}_j^{\mathrm{T}})\mathbf{A}^{\mathrm{T}}, \quad j = 1, 2, \ldots, J \\ \mathbf{X}_{i..} &= \mathbf{B}\mathrm{diag}(\mathbf{a}_i^{\mathrm{T}})\mathbf{C}^{\mathrm{T}}, \quad i = 1, 2, \ldots, I \end{aligned} \tag{9}$$

Theorem 4

Provided that $N \geq F$ and **A, B** and **C** all have rank F, the profiles of the underlying factors will be contained in the results obtained by SWATLD, i.e. regardless of some scaling indeterminacy, for every column of the underlying loading matrices $\mathring{\mathbf{A}}$, $\mathring{\mathbf{B}}$ and $\mathring{\mathbf{C}}$ an equivalent column always exists in the corresponding loading matrices **A, B** and **C** resolved by SWATLD.

Proof. From (8) and (9) one has

$$\mathbf{A}^{+}\mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}} = \mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}, \quad k = 1, 2, \ldots, K$$

Suppose

$$\mathbf{A} = (\mathbf{A}_F, \mathbf{A}_{N-F}), \qquad \mathbf{B} = (\mathbf{B}_F, \mathbf{B}_{N-F}), \qquad \mathrm{diag}(\mathbf{c}_k^{T}) = \begin{pmatrix} \mathrm{diag}(\mathbf{c}_{k,F}^{T}) & 0 \\ 0 & \mathrm{diag}(\mathbf{c}_{k,N-F}^{T}) \end{pmatrix}$$

where $\mathbf{A}_F$ of size $I \times F$ is a column full-rank matrix. Since one has

$$\mathbf{A}^{+} = \begin{pmatrix} \mathbf{A}_F^{+} - \mathbf{A}_F^{+}\mathbf{A}_{N-F}\mathbf{K}^{-1}\mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathbf{A}_F^{+} \\ \mathbf{K}^{-1}\mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathbf{A}_F^{+} \end{pmatrix}$$

$$\mathbf{K} = \mathbf{I} + \mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathbf{A}_F^{+}\mathbf{A}_{N-F}$$

hence

$$\begin{aligned}(\mathbf{A})^{+}\mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}} &= \begin{pmatrix} \mathbf{A}_F^{+} - \mathbf{A}_F^{+}\mathbf{A}_{N-F}\mathbf{K}_1^{-1}\mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathbf{A}_F^{+} \\ \mathbf{K}_1^{-1}\mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathbf{A}_F^{+} \end{pmatrix}(\mathbf{A}_F, \mathbf{A}_{N-F}) \\ &\quad \times \begin{pmatrix} \mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}}) & 0 \\ 0 & \mathrm{diag}(\mathbf{c}_{k,N-F}^{\mathrm{T}}) \end{pmatrix}(\mathbf{B}_F, \mathbf{B}_{N-F})^{\mathrm{T}} \\ &= \begin{pmatrix} \mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}}) & 0 \\ 0 & \mathrm{diag}(\mathbf{c}_{k,N-F}^{\mathrm{T}}) \end{pmatrix}(\mathbf{B}_F, \mathbf{B}_{N-F})^{\mathrm{T}}\end{aligned}$$

Consequently,

$$\begin{aligned}&(\mathbf{I} - \mathbf{A}_F^{+}\mathbf{A}_{N-F}\mathbf{K}_1^{-1}\mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}})\mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}})\mathbf{B}_F^{\mathrm{T}} \\ &\quad + \mathbf{A}_F^{+}(\mathbf{I} - \mathbf{A}_{N-F}\mathbf{K}_1^{-1}\mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}})\mathbf{A}_{N-F}\mathrm{diag}(\mathbf{c}_{k,N-F}^{\mathrm{T}})\mathbf{B}_{N-F}^{\mathrm{T}} = \mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}})\mathbf{B}_F^{\mathrm{T}} \\ &\mathbf{K}_1^{-1}\mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}})\mathbf{B}_F^{\mathrm{T}} + \mathbf{K}_1^{-1}\mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathbf{A}_F^{+}\mathbf{A}_{N-F}\mathrm{diag}(\mathbf{c}_{k,N-F}^{\mathrm{T}})\mathbf{B}_{N-F}^{\mathrm{T}} \\ &\quad = \mathrm{diag}(\mathbf{c}_{k,N-F}^{\mathrm{T}})\mathbf{B}_{N-F}^{\mathrm{T}}\end{aligned}$$

Thus we have

$$\begin{aligned}&\mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}})\mathbf{B}_F^{\mathrm{T}} + \mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathbf{A}_F^{+}\mathbf{A}_{N-F}\mathrm{diag}(\mathbf{c}_{k,N-F}^{\mathrm{T}})\mathbf{B}_{N-F}^{\mathrm{T}} \\ &\quad = \mathbf{K}_1\mathrm{diag}(\mathbf{c}_{k,N-F}^{\mathrm{T}})\mathbf{B}_{N-F}^{\mathrm{T}}\end{aligned}$$

This can be rewritten as

$$(\mathbf{K}_1 - \mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathbf{A}_F^{+}\mathbf{A}_{N-F})\mathrm{diag}(\mathbf{c}_{k,N-F}^{\mathrm{T}})\mathbf{B}_{N-F}^{\mathrm{T}} = \mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}})\mathbf{B}_F^{\mathrm{T}}$$

Finally we have the equation

$$\mathrm{diag}(\mathbf{c}_{k,N-F}^{\mathrm{T}})\mathbf{B}_{N-F}^{\mathrm{T}} = \mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}}\mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}})\mathbf{B}_F^{\mathrm{T}}$$

If $\mathbf{A}_{N-F} = 0$, the following equation obviously holds:

$$\mathbf{X}_{..k} = \mathbf{A}_F\mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}})\mathbf{B}_F^{\mathrm{T}}$$

If $\mathbf{A}_{N-F} \neq 0$,

$$\mathbf{X}_{..k} = \mathbf{A}_F \mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}})\mathbf{B}_F^{\mathrm{T}} + \mathbf{A}_{N-F}\mathrm{diag}(\mathbf{c}_{k,N-F}^{\mathrm{T}})\mathbf{B}_{N-F}^{\mathrm{T}}$$
$$= (\mathbf{A}_F + \mathbf{A}_{N-F}\mathbf{A}_{N-F}^{\mathrm{T}}(\mathbf{A}_F^{+})^{\mathrm{T}})\mathrm{diag}(\mathbf{c}_{k,F}^{\mathrm{T}})\mathbf{B}_F^{\mathrm{T}}$$

According to the definition in Appendix I,

$$\mathbf{X}_{..k} = \overset{\circ}{\mathbf{A}}_{I\times F}\mathrm{diag}(\overset{\circ}{\mathbf{c}}_k^{\mathrm{T}})\overset{\circ}{\mathbf{B}}_{J\times F}^{\mathrm{T}}$$

Suppose that the columns of $\mathbf{B}_F$ and $\mathring{\mathbf{B}}_{J\times F}$ are arranged according to the same criterion. Hence, according to the theorem of Kruskal [19,20], the F columns of loading matrix $\mathbf{B}_F$ should be the profiles in the second mode of the F underlying factors with physical meaning, i.e. $\mathbf{B}_F = \mathring{\mathbf{B}}_{J\times F}\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix. This conclusion implies that loading matrix $\mathbf{B}_F$ is also a column full-rank matrix. Therefore it can easily be demonstrated that the columns of loading matrices $\mathbf{A}_F$ and $\mathbf{C}_F$ should also be the corresponding profiles in the first and third modes respectively of the underlying factors with physical meaning. For any column of $\mathbf{A}$, $\mathbf{a}_i \neq 0$, there exist $F-1$ other columns $\mathbf{A}_{F-1}$ which satisfy $\mathrm{rank}([\mathbf{A}_{F-1},\mathbf{a}_i]) = F$. Thus the corresponding $\mathbf{b}_i$ should be the profile in the second mode of a certain underlying factor, which indicates that $\mathbf{a}_i$ and $\mathbf{c}_i$ are the respective profiles in the first and third modes of the certain underlying factor. In conclusion, the columns of the loading matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ obtained by SWATLD should be the profiles in the corresponding modes of the underlying factors, otherwise zero.

It should be noted that the above conclusion holds only when $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ all have rank F, which is generally satisfied when SWATLD is used.

4. SIMULATION STUDIES OF THE PROPERTY OF SWATLD BEING INSENSITIVE TO EXCESS FACTORS USED IN CALCULATION IN THE PRESENCE OF RANDOM NOISE

The property of SWATLD being insensitive to excess factors used in calculation has been proved under the assumption $\underline{\mathbf{E}} = 0$. Its performance in the presence of noise has to be demonstrated by simulation studies. Data arrays were simulated according to the scheme in Appendix II. The initialization and stopping criterion of SWATLD are also given in Appendix II.

Table I lists the results of SWATLD for a randomly simulated three-component data array of size $20 \times 20 \times 5$. The results show that SWALTD can always find the correct profiles of the underlying factors in the three modes within a reasonable number of iterations. The property of SWATLD being insensitive to excess factors used in calculation is also observed in the presence of noise. As expected, an increase in the number of excess factors has little effect on the quality of the results obtained by SWATLD. Even when an N that is three times the number of underlying factors ($N = 9$) is employed, the quality of the results of SWATLD shows no obvious deterioration, the resolved profiles of the underlying factors in the three modes being in good agreement with the true ones (Figure 1). With the noise level parameter a_{noise} varying from 0·002 to 0·02, the results of SWATLD with $N = 4$ or 5 are essentially the same as those when $N = 3$, the same as the number of underlying factors in the simulated system. These results have fully demonstrated that the attractive property of SWATLD being insensitive to excess factors used in calculation also holds in the presence of noise.

Table I. Influence of N on final results of SWATLD for a randomly simulated three-component data array with a_{noise} ranging from 0·002 to 0·02 (for each N value, five randomly initialized runs were performed)

			R_A[b]			R_B			R_C			
a_{noise}	N	IND[a]	1	2	3	1	2	3	1	2	3	IT[c]
0·002	3	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	26
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	23
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	25
	4	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	92
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	33
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	59
	5	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	270
		Min	0·9991	1·0000	0·9998	0·9994	1·0000	0·9998	0·9994	1·0000	0·9999	26
		Ave	0·9998	1·0000	1·0000	0·9999	1·0000	1·0000	0·9999	1·0000	1·0000	101
0·004	3	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	27
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	24
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	25
	4	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	612
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	58
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	430
	5	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	182
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	94
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	148
0·006	3	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	29
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	21
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	23
	4	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	125
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	27
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	74
	5	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	94
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	34
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	64
0·008	3	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	23
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	20
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	22
	4	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	93
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	30
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	50
	5	Max	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	88
		Min	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	60
		Ave	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	73
0·01	3	Max	0·9999	1·0000	1·0000	0·9999	1·0000	1·0000	1·0000	1·0000	1·0000	25
		Min	0·9999	1·0000	1·0000	0·9999	1·0000	1·0000	1·0000	1·0000	1·0000	15
		Ave	0·9999	1·0000	1·0000	0·9999	1·0000	1·0000	1·0000	1·0000	1·0000	22
	4	Max	0·9999	1·0000	1·0000	0·9999	1·0000	1·0000	1·0000	1·0000	1·0000	89
		Min	0·9999	1·0000	1·0000	0·9999	1·0000	1·0000	1·0000	1·0000	1·0000	43
		Ave	0·9999	1·0000	1·0000	0·9999	1·0000	1·0000	1·0000	1·0000	1·0000	69
	5	Max	0·9999	1·0000	1·0000	0·9999	1·0000	1·0000	1·0000	1·0000	1·0000	122
		Min	0·9999	1·0000	1·0000	0·9999	1·0000	1·0000	1·0000	1·0000	1·0000	43
		Ave	0·9999	1·0000	1·0000	0·9999	1·0000	1·0000	1·0000	1·0000	1·0000	70
0·02	3	Max	0·9997	0·9999	0·9998	0·9997	0·9999	0·9998	1·0000	0·9999	0·9998	25
		Min	0·9997	0·9999	0·9998	0·9997	0·9999	0·9998	1·0000	0·9999	0·9998	15
		Ave	0·9997	0·9999	0·9998	0·9997	0·9999	0·9998	1·0000	0·9999	0·9998	20
	4	Max	0·9997	0·9999	0·9998	0·9997	0·9999	0·9998	1·0000	0·9999	0·9999	113
		Min	0·9997	0·9999	0·9998	0·9996	0·9999	0·9998	1·0000	0·9999	0·9998	28

Table I. (*continued*)

a_{noise}	N	IND[a]	R_A[b] 1	2	3	R_B 1	2	3	R_C 1	2	3	IT[c]
		Ave	0·9997	0·9999	0·9998	0·9997	0·9999	0·9998	1·0000	0·9999	0·9999	56
	5	Max	0·9997	0·9999	0·9998	0·9997	0·9999	0·9998	1·0000	0·9999	0·9999	161
		Min	0·9997	0·9999	0·9998	0·9996	0·9999	0·9998	0·9999	0·9999	0·9998	32
		Ave	0·9997	0·9999	0·9998	0·9996	0·9999	0·9998	1·0000	0·9999	0·9998	76

[a] IND, Max, Min and Ave denote 'index', 'maximum', 'minimum' and 'average', respectively.
[b] $R_{\mathbf{A}}$, $R_{\mathbf{B}}$ and $R_{\mathbf{C}}$ are the relative coefficients between the resolved and the corresponding true profiles (for $N > F$, only three columns of the resolved loading matrices **A, B** and **C** are used to demonstrate the performance of SWATLD, since the rest represent noise).
[c] IT denotes the iterations required by SWATLD.

5. SOME DEDUCTIONS FROM THE EXPLANATIONS

The previous sections not only offer some mathematical explanations of the property of SWATLD being insensitive to excess factors used in calculation, but also give clues for further understanding other properties of SWATLD and guidelines for designing new trilinear decomposition algorithms.

5.1 In order to decompose data arrays into loading matrices with chemical meanings, SWATLD requires that the underlying loading matrices $\mathring{\mathbf{A}}$, $\mathring{\mathbf{B}}$ *and* $\mathring{\mathbf{C}}$ *should have equal rank, say F*

Assume $rank(\mathring{\mathbf{A}}) = rank(\mathring{\mathbf{B}}) = F$ and $rank(\mathring{\mathbf{C}}) = F - 1$. Suppose $S(\mathbf{C}) = 0$, $S(\mathbf{B}) = 0$ and $S(\mathbf{A}) = 0$ when SWATLD stops. Then

$$\mathbf{X}_{j\cdot} = (\mathbf{C}_{F-1} + \mathbf{C}_{F-1}\mathbf{C}_{N-F+1}^{\mathrm{T}}(\mathbf{C}_{F-1}^{+})^{\mathrm{T}})\mathrm{diag}(\mathbf{b}_{j,F-1}^{\mathrm{T}})\mathbf{A}_{I\times(F-1)}^{\mathrm{T}}, \quad j = 1, 2, \ldots, J$$

These equations are contradicted by the equations

$$\mathbf{X}_{j\cdot} = \mathring{\mathbf{C}}\mathrm{diag}(\mathring{\mathbf{b}}_j^{\mathrm{T}})\mathrm{A}_{I\times F}^{\mathrm{T}}, \quad j = 1, 2, \ldots, J$$

This contradiction results from the assumptions $S(\mathbf{C}) = 0$, $S(\mathbf{B}) = 0$ and $S(\mathbf{A}) = 0$. It means that if $\mathrm{rank}(\mathring{\mathbf{A}}) = \mathrm{rank}(\mathring{\mathbf{B}}) = F$ and $\mathrm{rank}(\mathring{\mathbf{C}}) = F - 1$, then $S(\mathbf{C}) \neq 0$, $S(\mathbf{B}) \neq 0$ and $S(\mathbf{A}) \neq 0$ when SWATLD stops.

The symmetry constraint is introduced by the three objective functions employed in SWATLD. Although the constraint limits the applications of SWATLD, its features of fast convergence and stable performance under different conditions are attractive in second-order linear calibration problems [18], since the symmetry constraint can be satisfied simply by preparing several more calibration samples.

5.2. Other possible iterative trilinear decomposition algorithms with the property of being insensitive to the number of factors chosen

From the above explanations it can be seen that the property of being insensitive to excess factors used in calculation originates from the objective functions adopted. Actually, it is unnecessary to employ three objective functions as in SWATLD if only the above property is concerned. The feature can be guaranteed simply by optimizing objective function (10) on condition that

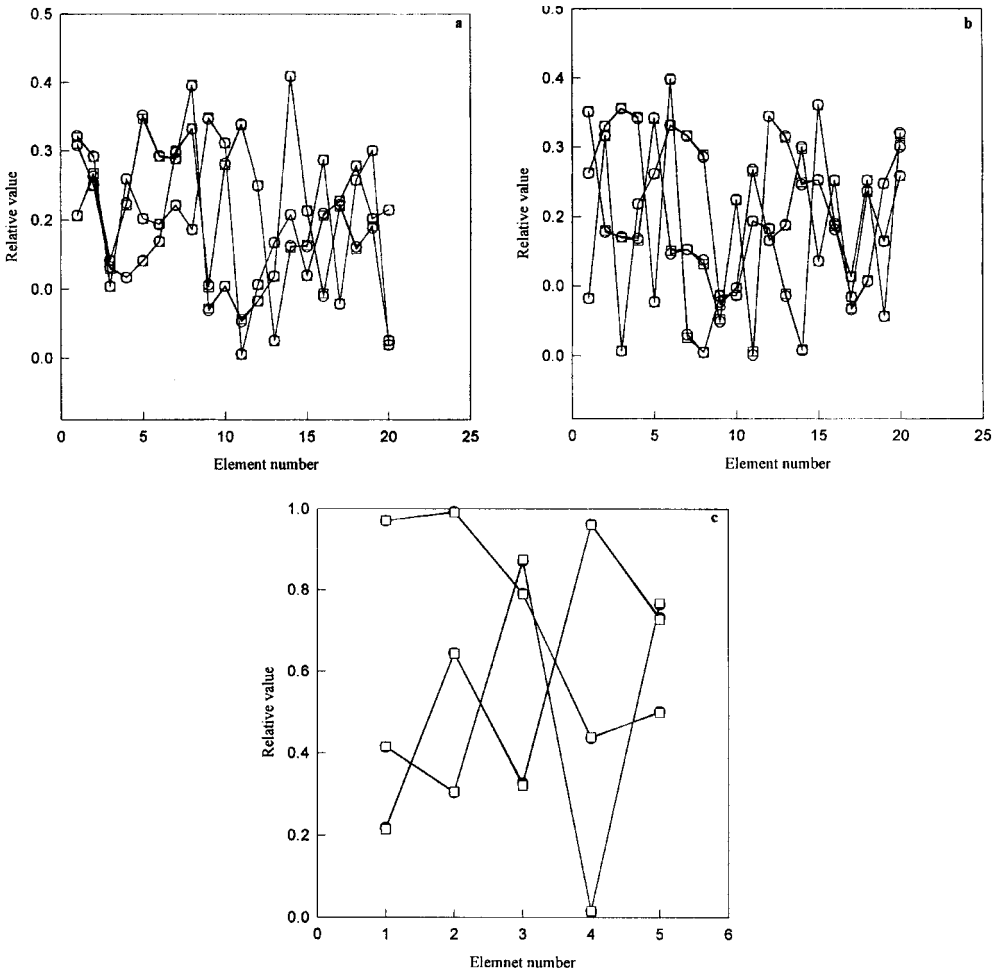

Figure 1. True (○) and resolved (□) loadings in three modes by SWATLD ($N = 9$) for a randomly simulated three-component data array of size $20 \times 20 \times 5$ with $a_{\text{noise}} = 0{\cdot}01$ (for the resolved loading matrices (a) **A**, (b) **B** and (c) **C**, only three of their columns have been depicted, since the rest represent noise).

$\text{rank}(\mathbf{A}) = \text{rank}_c(\sum_{k=1}^{K} \mathbf{X}_{..k})$ or optimizing objective function (11) on condition that $\text{rank}(\mathbf{B}) = \text{rank}_F(\sum_{k=1}^{K} \mathbf{X}_{..k})$:

$$S(\mathbf{C}) = \sum_{k=1}^{K} \left\| \mathbf{A}^{+}\mathbf{X}_{..k} - \text{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}} \right\|_{\mathrm{F}}^2 \tag{10}$$

$$S(\mathbf{C}) = \sum_{k=1}^{K} \left\| \mathbf{X}_{..k}(\mathbf{B}^{+})^{\mathrm{T}} - \mathbf{A}\text{diag}(\mathbf{c}_k^{\mathrm{T}}) \right\|_{\mathrm{F}}^2 \tag{11}$$

where $\text{rank}_c(\ )$ and $\text{rank}_r(\ )$ are the column and the row rank of a matrix respectively.

Using only one objective function can avoid the introduction of the symmetry constraint on the ranks of loading matrices $\mathring{\mathbf{A}}$, $\mathring{\mathbf{B}}$ and $\mathring{\mathbf{C}}$. However, there are still some extra assumptions on the loading matrices **A** and **B** obtained in order to draw the conclusion that the final results will not influenced by excess factors used in calculation. Furthermore, it should be pointed out that algorithms established on one of the above objective functions suffer from the same annoying problem of slow convergence as PARAFAC.

A much more theoretically favourable algorithm can be designed by alternately optimizing the following three objective functions (*i.e.* supposing **A** and **B** are known, minimize S(**C**) to obtain loading matrix **C**; with **A** and the new **C**, minimize S(**B**) to renew **B**; and then minimize S(**A**) with the latest **B** and **C** to obtain a new **A**):

$$S(\mathbf{C}) = \sum_{k=1}^{K} (2\left\|(\mathbf{X}_{..k} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}}))\mathbf{B}^{\mathrm{T}}\right\|_{\mathrm{F}}^2 + \lambda(\left\|\mathbf{A}^{+}\mathbf{X}_{..k} - \mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}\right\|_{\mathrm{F}}^2 + \left\|\mathbf{X}_{..k}(\mathbf{B}^{+})^{\mathrm{T}} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\right\|_{\mathrm{F}}^2)) \tag{12}$$

$$S(\mathbf{B}) = \sum_{k=1}^{K} (\left\|\mathbf{X}_{..k} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{A}^{+}\mathbf{X}_{..k} - \mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}\right\|_{\mathrm{F}}^2) \tag{13}$$

$$S(\mathbf{A}) = \sum_{k=1}^{K} (\|\mathbf{X}_{..k} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}_{..k}(\mathbf{B}^{+})^{\mathrm{T}} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\|_{\mathrm{F}}^2) \tag{14}$$

where λ is a scalar. It can easily be verified that without any extra constraints such as the symmetry constraint on the data array or assumptions on the loading matrices **A**, **B** and **C** calculated, the above algorithm has the feature of being insensitive to excess factors used in calculation. Its potential applications will appear in the near future.

6. CONCLUSIONS

The property of SWATLD being insensitive to excess factors used in calculation is rooted in the objective functions utilized. As far as this property is concerned, it is unnecessary to employ three different objective functions, which will introduce a symmetry constraint on the data array analysed. Under some mild conditions the chemical meaning of the results can be guaranteed through optimizing only one appropriate objective function such as

$$S(\mathbf{C}) = \sum_{k=1}^{K} \|\mathbf{A}^{+}\mathbf{X}_{..k} - \mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}\|_{\mathrm{F}}^2 \quad \text{or}$$

$$S(\mathbf{C}) = \sum_{k=1}^{K} \|\mathbf{X}_{..k}(\mathbf{B}^{+})^{\mathrm{T}} - \mathbf{A}\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})\|_{\mathrm{F}}^2$$

even if $N > F$. At the same time it can also avoid the introduction of the symmetry constraint. However, preliminary studies showed that algorithms based on just one objective function have the problem of slow convergence. As long as $N \geq F$, the algorithm based on objective functions (12), (13) and (14) can always guarantee the chemical meaning of the results, without any extra constraints on the data array or assumptions on the loading matrices **A**, **B** and **C** obtained.

ACKNOWLEDGEMENTS

The authors would like to thank the National Natural Science Foundation of China for financial support (grant 29735150).

APPENDIX I. NOMENCLATURE

x_{ijk}	the ijkth element of the three-way array $\underline{\mathbf{X}}$
I, J, K	the dimensions of different modes in three-way arrays
F	the number of underlying factors, i.e. the total number of detectable species, including the components of interest and interference(s) as well as background
N	the number of factors used in calculation
$\mathring{a}_{if}$, $\mathring{b}_{if}$, and $\mathring{c}_{kf}$	the ifth, jfth and kfth elements of the three underlying loading matrices $\mathring{\mathbf{A}}_{I\times F}$, $\mathring{\mathbf{B}}_{J\times F}$ and $\mathring{\mathbf{C}}_{K\times F}$ in data array $\underline{\mathbf{X}}$ respectively (in this paper, $\mathring{\mathbf{A}}_{I\times F}$, $\mathring{\mathbf{B}}_{J\times F}$ and $\mathring{\mathbf{C}}_{K\times F}$ are simply written as $\mathring{\mathbf{A}}$, $\mathring{\mathbf{B}}$ and $\mathring{\mathbf{C}}$ respectively)
$\mathbf{A}_{I\times N}$, $\mathbf{B}_{J\times N}$, $\mathbf{C}_{K\times N}$	the three resolved loading matrices of $\underline{\mathbf{X}}$ with dimensions $I\times N$, $J\times N$ and $K\times N$ respectively (for simplicity, in this paper, $\mathbf{A}_{I\times N}$, $\mathbf{B}_{J\times N}$ and $\mathbf{C}_{K\times N}$ are represented by **A, B** and **C** respectively)
$\mathbf{A}^{+}$	the Moore–Penrose generalized inverse of matrix **A**
$\mathbf{X}_{..k}=\mathring{\mathbf{A}}\mathrm{diag}\,(\mathring{\mathbf{c}}_k^{\mathrm{T}})\,\mathring{\mathbf{B}}^{\mathrm{T}}$	the kth frontal slice of the three-way array $\underline{\mathbf{X}}$
$\mathbf{X}_{.k.}=\mathring{\mathbf{C}}\mathrm{diag}\,(\mathring{\mathbf{b}}_j^{\mathrm{T}})\,\mathring{\mathbf{A}}^{\mathrm{T}}$	the jth lateral slice of the three-way array $\underline{\mathbf{X}}$
$\mathbf{X}_{i..}=\mathring{\mathbf{B}}\mathrm{diag}\,(\mathring{\mathbf{a}}_i^{\mathrm{T}})\,\mathring{\mathbf{A}}^{\mathrm{T}}$	the ith horizontal slice of the three-way array $\underline{\mathbf{X}}$
$\mathrm{diag}(\mathring{\mathbf{c}}_k^{\mathrm{T}})$, $\mathrm{diag}(\mathring{\mathbf{a}}_i^{\mathrm{T}})$ and $(\mathring{\mathbf{b}}_j^{\mathrm{T}})$	diagonal matrices with elements equal to the kth, ith and jth rows of $\mathring{\mathbf{C}}$, $\mathring{\mathbf{A}}$ and $\mathring{\mathbf{B}}$ respectively
$\mathrm{diag}(\mathbf{c}_k^{\mathrm{T}})$, $\mathrm{diag}(\mathbf{a}_i^{\mathrm{T}})$, $\mathrm{diag}(\mathbf{b}_j^{\mathrm{T}})$	diagonal matrices with elements equal to the kth, ith and jth rows of **C, A** and **B** respectively
$\mathrm{diagm}(\mathbf{A}^{\mathrm{T}}\mathbf{A})$, $\mathrm{diagm}(\mathbf{B}^{\mathrm{T}}\mathbf{B})$, $\mathrm{diagm}(\mathbf{C}^{\mathrm{T}}\mathbf{C})$	vectors with elements equal to the diagonal elements of matrices $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, $\mathbf{B}^{\mathrm{T}}\mathbf{B}$ and $\mathbf{C}^{\mathrm{T}}\mathbf{C}$ respectively

APPENDIX II

II. 1. The scheme for simulating data arrays

$$\mathring{\mathbf{A}}=\mathbf{rand}(I,F)$$

$$\mathring{\mathbf{B}}=\mathbf{rand}(J,F)$$

$$\mathring{\mathbf{C}}=\mathbf{rand}(K,F)$$

$$\mathbf{X}^{p}_{..k}=\mathring{\mathbf{A}}\mathrm{diag}(\mathring{\mathbf{c}}_k^{\mathrm{T}})\mathring{\mathbf{B}}^{\mathrm{T}}$$

$$\mathbf{E}_{..k}=\mathrm{Max}(\mathbf{X}^{p}_{..k})\times\mathbf{randn}(I,J)\times a_{\mathrm{noise}},\quad k=1,2,\ldots,K$$

$$\mathbf{X}_{..k}=\mathbf{X}^{p}_{..k}+\mathbf{E}_{..k},\quad k=1,2,\ldots,K$$

$$\underline{\mathbf{X}}=\{\mathbf{X}_{..k}\}$$

where a_{noise} is a scalar controlling the noise level added, **rand**(I,F) is an $I\times F$ matrix with random elements taking values in the range (0, 1), **randn**(I,J) is an $I\times J$ matrix with random entries chosen

from a normal distribution with mean zero and variance one, and $\mathrm{Max}(\mathbf{X}^p_{..k})$ is the maximal element of matrix $\mathbf{X}^p_{..k}$.

II. 2. The initialization and stopping criterion of SWATLD

Random initialization was carried out to start the iterative optimization procedure of SWATLD. The optimization procedure is terminated when the following criterion reaches a certain threshold ε ($\varepsilon = 1 \times 10^{-6}$ in the present paper)—a maximal iteration number of 3000 is adopted to avoid possible unduly slow convergence

$$\mathrm{SSR}^{(m)} = \sum_{k=1}^{K} \left\| \mathbf{X}_{..k} - \mathbf{A}^{(m)} \mathrm{diag}\left((\mathbf{c}^{(m)})^{\mathrm{T}}_k \right) \mathbf{B}^{(m)^{\mathrm{T}}} \right\|^2_{\mathrm{F}}$$

$$\left| \frac{\mathrm{SSR}^{(m)} - \mathrm{SSR}^{(m-1)}}{\mathrm{SSR}^{(m-1)}} \right| \leq \varepsilon$$

where SSR is the residual sum of squares and m is the current iteration number.

REFERENCES

1. Liang Y-Z. *White, Grey and Black Multi-component Systems and Their Chemometric Algorithms*. Hunan Publishing House of Science and Technology: Changsha, 1996.
2. Sanchez E, Kowalski BR. *J. Chemometrics* 1990; **4**: 29.
3. Harshman RA. *UCLA Working Papers Phonet.* 1970; **16**: 1.
4. Carroll JD, Chang J. *Psychometrika* 1970; **35**: 283.
5. Burdick DS, Tu XM, Mcgown LB, Millican DW. *J. Chemometrics* 1990; **4**: 15.
6. Li S, Hamilton JC, Gemperline PJ. *Anal. Chem.* 1992; **64**: 599.
7. Paatero P. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 223.
8. Faber NM, Buydens LMC, Kateman G. *J. Chemometrics* 1994; **8**: 141.
9. Faber NM, Buydens LMC, Kateman G. *J. Chemometrics* 1994; **8**: 181.
10. Faber NM, Buydens LMC, Kateman G. *J. Chemometrics* 1994; **8**: 273.
11. Bro R. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 149.
12. Smilde AK. *Chemometrics Intell. Lab. Syst.* 1992; **5**: 143.
13. Bro R, Heimdal H. *Chemometrics Intell. Lab. Syst.* 1996; **34**: 85.
14. Nørgaard L. *Talanta* 1995; **42**: 1305.
15. Windig W, Antalek B. *Chemometrics Intell. Lab. Syst.* 1999; **46**: 207.
16. Smilde AK, Doornbos DA. *J. Chemometrics* 1991; **5**: 345.
17. Wu H-L, Shibukawa M, Oguma K. *J. Chemometrics* 1998; **12**: 1.
18. Chen Z-P, Wu H-L, Jiang J-H, Li Y, Yu R-Q. *Chemometrics Intell. Lab. Syst.* 2000; **52**: 75.
19. Harshman RA, Lundy ME. *Research Methods for Multimode Data Analysis*. Praeger: New York, 1984.
20. Kruskal JB. *Linear Algebra Appl.* 1977; **18**: 95.
21. Kruskal JB. *Psychometrika* 1976; **41**: 281.
22. Cline RE. *J. Soc. Ind. Appl. Math.* 1964; **12**: 588.

